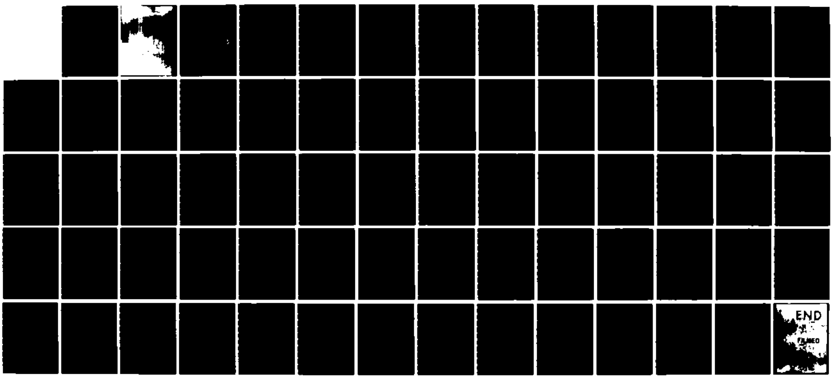
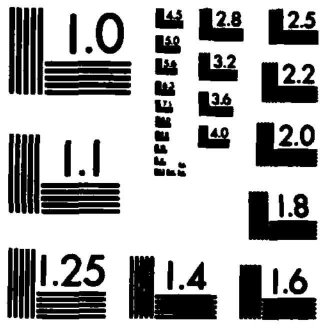
AD-A145 696   AN INTRODUCTION TO MULTILINEAR FORMULA SCORE THEORY(U)   1/1
ILLINOIS UNIV AT URBANA MODEL BASED MEASUREMENT LAB
M V LEVINE 28 AUG 84 MEASUREMENT SER-84-4

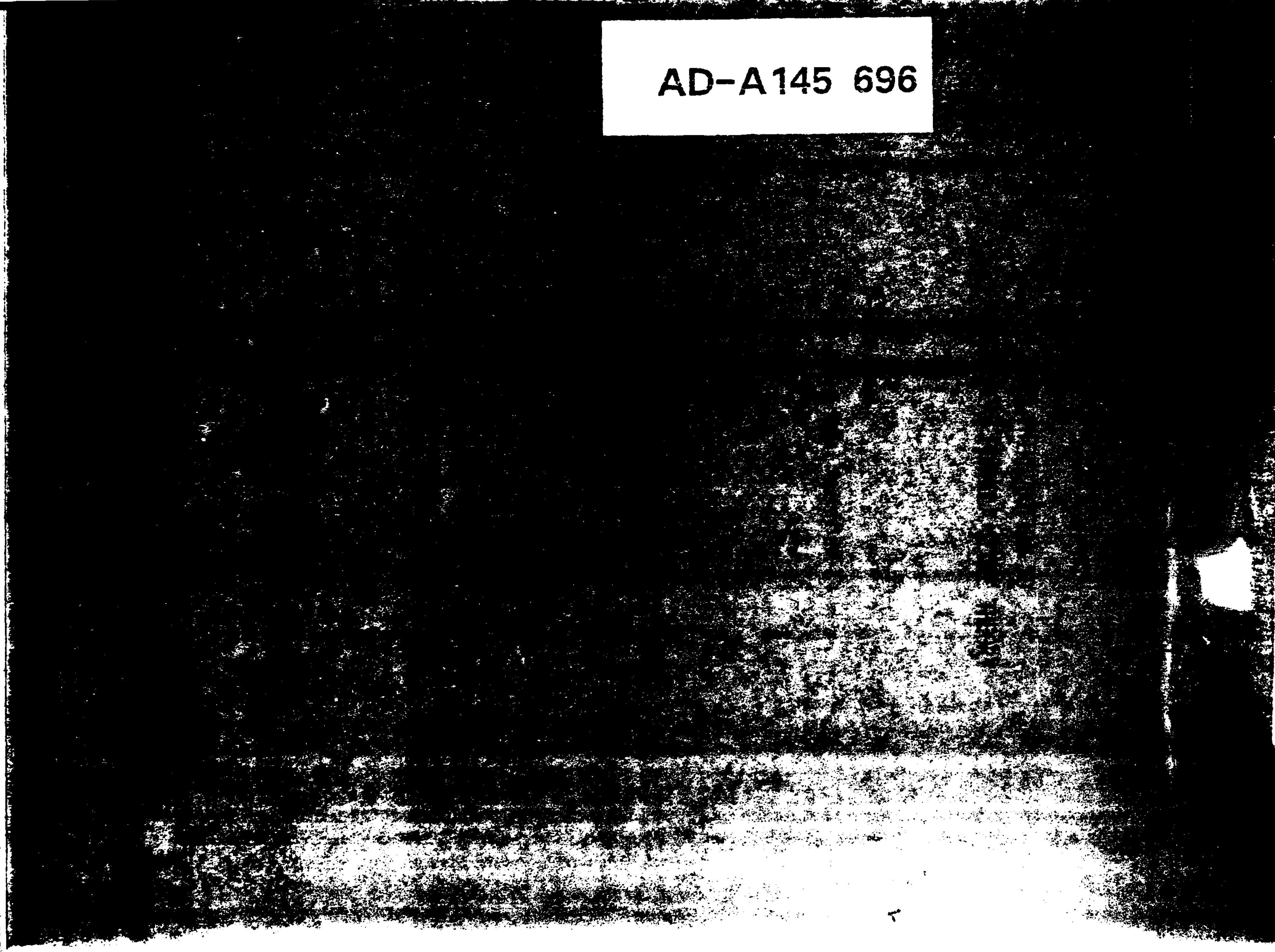
UNCLASSIFIED   N00014-79-C-0752   F/G 12/1   NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A145 696



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Measurement/Series 84-4	2. GOVT ACCESSION NO. A145696
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): An Introduction to Multilinear Formula Score Theory	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Michael V. Levine	8. CONTRACT OR GRANT NUMBER(s): N00014-79C-0752 N00014-83K-0397
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Model-Based Measurement Laboratory University of Illinois Champaign, IL 61820	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61153N RR042-04 NR 154-445 NR 150-518
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs Office of Naval Research (Code 442PT) Arlington, VA 22217	12. REPORT DATE: 28 August 1984
	13. NUMBER OF PAGES: 55
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This is an annotated, edited version of a 1982 survey of formula score theory. Sections two and three provide an outline of the basic theory. The remaining sections deal with applications.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Latent trait theory, item response theory, formula score theory, local independence test, ability distribution estimation, density estimation, item bias, item drift, multidimension items, multidimensional ability distributions, unidimensionality, test security, consistent estimation, maximum likelihood estimation, regression function.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Formula score theory (FST) associates each multiple choice test with a linear operator and expresses all of the real functions of item response theory as linear combinations of the operator's eigenfuncitons. Hard measurement problems can then often be reformulated as easier, standard mathematical problems. For example, the problem of estimating ability distributions from sequences of item responses can be reformulated as maximizing a convex index of goodness of fit defined on a convex set. A major simplification of several theoretical

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

problems has been obtained because the linear mathematics used by the theory has a well-developed generalization to problems involving many variables. For example, a battery of tests measuring several related variables and one test measuring one trait can be analyzed with essentially the same theory.

An elementary outline of the basic theory is presented along with a discussion of several illustrative applications.

Preface

This report is an edited, annotated version of a paper presented to the Office of Naval Research Contractor's Conference in 1982. Comments have been inserted to bring the report up-to-date and to make it easier to read. The original paper will eventually be distributed in the conference proceedings under the title "The Trait in Latent Trait Theory." This version is being released now to serve as an introduction to the theory.

AN INTRODUCTION TO MULTILINEAR FORMULA SCORE THEORY

Introduction

A list of potential applications of the theory and an overview of the report is given in this section. Since 1982 work has begun on two additional applications. The theory provides a formula for calculating an adaptive test score that has the same conditional expected value as the number-right score for a conventional test. The theory also is being used to formulate a way to estimate the parameters of adaptive test items without interrupting operational testing.

Just what is being quantified in a latent trait or item response theory analysis of a mental test? A good theory may shed light on the following practical problems.

1. Deciding whether two tests measure the same trait or traits;
2. Analyzing the relative contributions of a pair of traits or abilities to test performance;
3. Detecting "functional" changes in items including those caused by security problems, mode of administration changes and changes in the familiarity with the concepts supporting the item in the population being tested;
4. Determining the adequacy of an "item response function," i.e., a specific mathematical formula relating performance to ability;
5. Discovering the shape of the item response functions including multidimensional item response functions;
6. Quantifying the magnitude and reliability of violations of the principal assumption of latent trait theory, "local independence;"
7. Modelling item responses (such as omitting or changing answers) that fail to be locally independent.

Some theoretical results bearing on these problems will be outlined. The central problem for the new theory is to represent traits, abilities or achievements, and their distributions.

The theory will first be motivated by an informal discussion of some of its applications. After the basic theory is presented, the discussion of applications will be resumed. Projected work and work in progress is also described.

AN INTRODUCTION TO
MULTILINEAR FORMULA SCORE THEORY

Section One: Motivation

Three Practical Problems

This section should be skipped on first reading. It was written to stimulate interest in theory by showing its relevance to some important applied problems. Unfortunately, it is hard to read and out-of-date since our new parameter estimation programs permit us to implement the applications with much smaller sample sizes. It contains nothing that is needed to understand the sections that follow.

Three important measurement problems will be used to motivate the theory. An attempt has been made to keep this paper self contained. However, this section, which may be skipped or skimmed, requires familiarity with two latent trait theory terms, "item response function" and "local independence." They are now defined for the special situations considered in this section and redefined in later sections where more generality is needed.

The item response function (also called the item characteristic curve, ICC, and conditional response function) is the (conditional) probability of sampling an examinee correctly answering the item from the subpopulation of all examinees at a particular ability level. Thus, if ability is unidimensional, then the item response function for the i^{th} item on a test is the real function P_i

$P_i(t)$ = the probability of a correct response to item i from an examinee sampled from all those with ability = t .

A pair of items, say the i^{th} and j^{th} , are said to be locally independent if they are independent in subpopulations having no variation in ability, i.e., if for all ability levels t ,

$$\text{Prob}\{\text{items } i \text{ and } j \text{ are both correct} \mid \text{ability} = t\}$$

equals the product of the item response functions

$$P_i(t)P_j(t) .$$

Three hard measurement problems inevitably arise in the maintenance of testing programs that attempt to give more or less the same test year after year to a large number of people. Examples of such programs are the military

entrance and placement programs, college entrance exams, graduate and professional school admissions exams, high school and grade school aptitude and achievement tests, and interest measures such as job satisfaction scales used in industrial settings. Some of those programs test over a hundred thousand examinees every year.

The three problems —— functional item change, item response function adequacy and local independence failure —— are now described.

Functional item change: An item may function differently, i.e., have different psychometric properties in two test administrations. For example, a vocabulary item requiring exposure to political terminology may seem relatively easy in a presidential election year. School curriculum changes, security problems, method of administration change, improper coaching, and item format change also may result in functional item change. The principal question here is to determine to what extent, if at all, an item has functionally changed.

Item response function adequacy: Many mathematical formulas have been proposed to represent item response functions. Psychological arguments have been used to challenge the correctness of each, usually over an ability range. For example, monotonic curves have been criticized for giving an incorrect representation over a low ability range because very low ability examinees may perform somewhat better than examinees just bright enough to be misled by item construction tricks. Curves that asymptote to one have been criticized because they contain no provision for the careless mistakes of very bright examinees answering items beneath their ability level. Of course, with the huge samples of examinees currently available, virtually any guess or estimate of the population IRF can be rejected. The goal in adequacy

problems is to determine whether a proposed curve is "adequate," i.e., close enough to the population IRF over an ability range to be acceptable in a specific application.

Local independence failure: Psychological reasoning or data analysis can sometimes lead one to suspect that the local independence assumption has been seriously violated. For example, a pair of reading comprehension items referring to the same reading passage may be, to an unacceptable degree, measuring familiarity with the content of the passage. An example arising in an empirical item bias study is described later in this section. One application of the theory being developed is to determine the magnitude and reliability of suspected local independence failures.

After some preliminaries, the three problems will be considered separately. Table I summarizes the discussion. It may be helpful for the reader to refer back to it from time-to-time.

TABLE I Summary of Formula Score Theory Analysis of Three Basic Measurement Problems

	Problem	Hypothesis	Population Parameter	Distribution of Test Statistic $\hat{\eta}$
1.	Functional Change	$L^* = P$	$\eta = \int_a^b (P-L)^2$	Quadratic function of normal variables; Non-central case.
2.	IRF Adequacy	$L = P$	$\eta = \int_a^b (P-L)^2$	Quadratic function of normal variables; Central case.
3.	Local Independence	$P_1P_2 = P$ $P_1 = L_1$ $P_2 = L_2$	$\eta = \int_a^b (P-L_1L_2)^2$	Quadratic function of normal variables; Central case.

It will be shown that all the questions can be reduced to a single question about two curves or functions: How close is a specified (either by a formula or table) function $L(\cdot)$ to an incompletely specified function $P(\cdot)$, the population item response function. The problems are hard to answer because abilities are not observed, only estimated.

Consider the conceptually simplest problem type, item response function adequacy. A "three parameter logistic function" L has been estimated and offered as a representation of the "true," i.e., population item response function P. The psychometrician is concerned about the monotonicity of P and suspects that L fits P poorly over, say, the ability range $-3 \leq \theta \leq -2$. He wishes to determine how far apart P and L are over this range.

An intuitive and commonly used measure of the distance between two functions is the generalization of Euclidean distance given by the root mean square of function values. In this spirit, an attempt will be made to compute a point estimate and confidence interval for

$$\eta = \int_{-3}^{-2} [P(t) - L(t)]^2 dt .$$

The interval $[-3,-2]$ in the definition of η is arbitrary. The hypothesis being tested and the sample of examinees available for testing will generally suggest a different center and width of the "supporting" interval. Short intervals give more specific information about the difference between P and L. Very short intervals give estimates of η with a large sampling error.

The results of this section are made possible by an elementary equation which is valid at each point t where ability densities are continuous.

$$(1)\quad [P(t) - L(t)]f(t) = f^{+}(t)[1 - L(t)]\bar{P} + f^{-}(t)L(t)\bar{Q}$$

In this equation P and L are as before. The density for the ability θ in the general population of examinees is denoted by f . f^{+} is the ability density in the subpopulation of examinees passing the target item; f^{-} is the conditional density in the failure subpopulation. $\bar{P} = 1 - \bar{Q}$ is the proportion of item passers.

This equation is important because it permits one to evaluate adequacy questions without estimating abilities. It is necessary to do so because for a test of fixed length any estimate of ability has a substantial standard error, a bound for which can be computed by routine methods. On the other hand, subject to technical qualifications treated at length below, an arbitrarily accurate estimate of the distribution of abilities can be obtained with a sufficiently large sample of examinees, and test administrations of over 1,000,000 examinees are no longer uncommon. In later sections, consistent estimates of ability densities are discussed.

In view of the very large sample sizes, f will be regarded as known. The effects of small errors in specifying f on the sampling distribution of our estimate of η has not been worked out at this time.

The quantity $\bar{P}$ on the right hand side can either be computed as $\int L(t)f(t)dt$ when the hypothesis $P = L$ is being evaluated or estimated as the sample proportion of item passers.

In most applications, only moderately large samples are available for estimating the conditional densities f^+ and f^-. Using Equation (1), one can write

$$(2) \quad \eta = \int_a^b \{f^+(t)[1-L(t)]\bar{P} + f^-(t)L(t)\bar{Q}\}^2 W(t)dt$$

where the weight function $W(t)$ is $1/[f(t)]^2$. Upon substituting estimates $\hat{f}^+(\cdot)$ and $\hat{f}^-(\cdot)$ for $f^+(\cdot)$ and $f^-(\cdot)$, one obtains a statistic $\hat{\eta}$,

$$(3) \quad \hat{\eta} = \int_a^b \{\hat{f}^+(t)[1-L(t)]\bar{P} + \hat{f}^-(t)L(t)\bar{Q}\}^2 W(t)dt$$

that can be used to evaluate IRF adequacy. It will be seen that $\hat{\eta}$ has a tractable sampling distribution.

To motivate some theoretical developments on density representation and estimation in the next section, suppose the conditional densities could be represented in the form

$$(4) \quad f^+(\theta) = \sum_{j=1}^{J} \alpha_j^+ h_j(\theta)$$

$$(5) \quad f^-(\theta) = \sum_{j=1}^{J} \alpha_j^+ h_j(\theta)$$

for known functions $h_1, h_2, \ldots, h_J$ and constants α_j^+, α_j^-.

Then, after the indicated integration in (2) is carried out, η has the particularly simple form

$$(6) \quad \eta = \underset{\sim}{\alpha} Q \underset{\sim}{\alpha}^T$$

where $\underset{\sim}{\alpha}$ is the vector $\langle\alpha_1^+,\ldots,\alpha_J^+,\alpha_1^-,\ldots,\alpha_J^-\rangle$ and Q is a matrix of numbers that can be calculated prior to data collection. The entries in Q are obtained by substituting (4) and (5) into (2), expanding the product and numerically calculating the integral of the product of the specified functions. It is easily verified that Q is symmetric, positive semidefinite.

Such a representation has been derived. The functions h_j are derived from a priori considerations given in the next section. The number of them, J , turns out to be acceptably small — between 4 and 8 — for the tests already analyzed.

Consistent, unbiased estimates for the vector of constants are described in the following sections. With them, one obtains estimated densities

$$\hat{f}^+(\cdot) = \sum_{j=1}^{J} \hat{\alpha}_j^+ h_j(\cdot)$$

$$\hat{f}^-(\cdot) = \sum_{j=1}^{J} \hat{\alpha}_j^- h_j(\cdot) \ .$$

Here $\hat{\alpha}_j^+$ and $\hat{\alpha}_j^-$ are estimates of the corresponding constants.

The vector of estimates $\hat{\underset{\sim}{\alpha}}$ will be seen to be multivariate normal, at least asymptotically. The hypothesis P = L permits one to calculate, prior to data collection, the covariance matrix of the estimates and derive the distribution of

$$\hat{\eta} = \hat{\underset{\sim}{\alpha}} Q \hat{\underset{\sim}{\alpha}}^T$$

Random variables of form $\hat{\alpha}Q\hat{\alpha}^T$ where $\hat{\alpha}$ is multivariate normal and Q positive, semidefinite, generalize the χ^2 family of random variables.

In the "central case," the statistic has the same distribution as the sum of squares of several independent normal variables with zero mean and not necessarily equal variances. In the "non-central case," the means may be unequal. The asymptotic normality and the hypothesis $P = L$ make the central case appropriate. A numerical algorithm has been developed for computing the cdf $F(x) = \text{Prob}\{\hat{\eta} < x\}$ and determining the probability of observing an $\hat{\eta}$ equal to the sample value or larger under the hypothesis, $\eta = 0$, i.e., $P = L$ (Williams, 1984). For a review of alternative algorithms see Kotz et. al., 1967a,b. Technical details on the derivation and distribution of $\hat{\eta}$ are in Levine, 1983.

The above approach can be used to attack functional item change questions. In the treatment of adequacy, the distribution of $\hat{\eta}$ was derived under the hypothesis $P = L$. In studying change, the discrepancy between P and L is measured under they hypothesis that $P = L^*$ where L^* is some specified function other than L. Suppose, for example, an item response function has been carefully measured using a very large sample and that years of successful experience with the item were consistent with the IRF used to represent it. The hypothesis, $P = L^*$ has considerable support. However, after a security problem comes to light, a re-estimation with a smaller sample gives a function $L \neq L^*$. Further suppose only low ability examinees are motivated to exploit the security problem. Under the hypothesis $P = L^*$, how large is the squared difference between P and L expected to be over the low ability range? By a generalization of the arguments described above, the distribution of $\hat{\eta}$ can be derived. It turns out to be a quadratic function of normal variables, non-central case. Formulas for the variances and non-centrality

parameters are in Levine, 1983.

The method of this section suggests a way to quantify suspected departures from local independence. For example, in an item bias study in progress, a vocabulary item using the word "hurl" was found to be severely and reliably biased against sixth grade girls and in favor of sixth grade boys in two independent samples of 4,000 and 2,000 children. It seems likely that performance on another item also using a word favored by baseball writers would agree more with the hurl item score than predicted by the local independence assumption of latent trait theory. To analyze the causes and consequences of bias it would be valuable to have a method for measuring the magnitude and reliability of local independence violations over specified ability ranges.

To test for local independence, two suspect items may be (conjunctively) paired to form a complex item, an item that is scored correct if both component items are correct and incorrect otherwise. If the items are locally independent, then the item characteristic curve of the complex item will be the product of item characteristic curves of the component items. Evidence for a violation of local independence would be small $\eta_1 = \int_a^b [P_1(t) - L_1(t)]^2 dt$ and $\eta_2 = \int_a^b [P_2(t) - L_2(t)]^2 dt$ but large $\eta_{12} = \int_a^b [P_{1\&2}(t) - L_1(t)L_2(t)]^2 dt$.

In all the above examples, densities and conditional densities were represented as linear combinations of a finite set of known functions. It will be shown in later sections that <u>every</u> density is, in a sense soon to be made precise, equivalent to exactly one of these linear combinations. Every test will be shown to have associated with it a unique "canonical space" or vector space of functions equivalent to densities. It is hoped that this discussion shows that a theory for density representation and estimation

can be used to attack fundamental issues in psychological measurement. A new theory is outlined in the next section. Some further applications to basic scientific questions follow.

AN INTRODUCTION TO MULTILINEAR FORMULA SCORE THEORY

Section Two: The Canonical Space and Equivalent Ability Distributions

The statistics X_j described in this section and the next were being used in 1982 to estimate "coordinates." They remain important for the theory because they show that the "identifiable part" of a density can be estimated. However, much more efficient coordinate estimation strategies are now available for applications.

In the preceding section a relation was noted between several hard, important substantive psychological issues and the more routine methodological problem of density estimation. The approach of the previous section required a representation of ability densities as finite linear combinations of known functions and multivariate normal estimates of the coefficients in the combinations.

In this section an a priori derivation of the representation is reviewed on a very informal level. A somewhat more formal presentation of the derivation is outlined in the next section. This approach to psychometric problems will be called "formula scoring" or "formula score theory"[1] and abbreviated FS and FST.

The analysis is organized about three fundamental theoretical issues and methodological problems.

1. Ability Distribution Equivalence: Which, if any, pairs of fundamentally different ability distributions lead to exactly the same probability distributions on the only observables in testing, the item scores? What are necessary and sufficient conditions for two distributions to be equivalent (in the sense of making the same predictions)? What statements about the distribution are, in the technical, foundations of measurement sense of the term, meaningful?

[1]The term "multilinear formula score theory" is now being used to distinguish the theory from earlier work by other authors on linear combinations of item scores.

2. <u>Ability Distribution Representation</u>: Find a decomposition of an arbitrary ability density into two uniquely determined parts

$$f(\cdot) = f_0(\cdot) + f^*(\cdot)$$

such that two densities f_1 and f_2 are equivalent if and only if $f_1^* = f_2^*$. Find a finite dimensional parameterization of the "identifiable part" f^* of the ability density f.

3. <u>Ability Distribution Identification and Estimation</u>: Show that the "identifiable part" of the ability density is identifiable in the sense that a consistent estimate of $f^*(t)$ exists for each t. Construct an estimator.

The results of this section are derived from a version of latent trait theory informally presented now and somewhat more formally stated in the next section. The major random variables of the latent trait model are abilities θ and item scores $u_1, u_2, \ldots, u_n$. Examinees are considered to be randomly sampled from an infinite population of examinees. The "points" in the probability space of the basic latent model are examinees. Each examinee has a specific ability and (non-random) vector of item responses. Abilities and item responses are non-trivial random variables only because examinees are sampled. Item responses are assumed to be "locally independent", i.e., independent in the subpopulations of examinees defined by conditioning upon ability. Although it may not be obvious at first reading, this conceptualization of latent trait theory is compatible with the usual treatment of item responses as independent binomial random variables with success probabilities that are functionally dependent on abilities, provided no item is ever administered two times to the same examinee.

To attack the problems of identifying, representing and estimating ability distributions from a foundations of measurement point of view, the set of all statistics for a test is studied. Since only the item scores u_i are observed and since examinees work independently of one another, the set of all statistics is simply the set of number-valued functions of the item score random variables. Moreover this set can be shown to be a finite dimensional vector space.

An important tool for studying ability distributions in formula score theory is the canonical space of a test, formulated by referring to regression functions. The regression function of a statistic S is the real function

$$R_S(t) = E[S|\theta = t].$$

The canonical space (CS) of a test is the vector space of all regression functions. It is easily shown to be finite dimensional. In fact, in many FST applications it has been possible to treat it as a vector space of low (less than 8) dimensionality. (See Section 4 for further discussion of CS dimensionality.)

Before proceeding, several assumptions commonly used in FST are listed. First, the "item characteristic curves" $P_i = R_{u_i}$

$$\begin{aligned} P_i(t) &= \text{Probability that item } i \text{ is answered correctly given an examinee with ability equal to } t \text{ has been sampled} \\ &= E(u_i|\theta = t) \end{aligned}$$

are assumed to be continuous. In addition, all abilities are assumed to lie in a closed bounded interval I. The assumption of continuous P_i is restrictive and may have to be dropped for some

applications. The assumption of bounded abilities, on the other hand, results in no loss of generality because any latent trait model can be reformulated by routine methods as an isomorphic model with bounded abilities. These assumptions together imply that the canonical space consists of continuous functions on the interval I .

A major result of FST is that two densities are equivalent in the sense of (1.) if they have the same projection into the canonical space . Thus if some J functions $h_1, h_2, \ldots, h_J$ form a basis for the CS and if

$$\int_I h_j(t) f_1(t)dt = \int_I h_j(t) f_2(t)dt \qquad j=1,2,\ldots,J$$

then there is no objective way to choose between f_1 and f_2 .

By an "objective way to choose between f_1 and f_2" is meant a method of using the observables (the item scores) to decide which of f_1 or f_2 is more nearly correct. This is impossible because it can be proven that every statistic has the same probability distribution when f_1 is correct as when f_2 is correct.

This fact leads to a useful representation of densities. An arbitrary density f can be represented uniquely as

$$f(\cdot) = f_0(\cdot) + \sum_{j=1}^{J} \alpha_j h_j(\cdot)$$

$$= f_0(\cdot) + f^*(\cdot)$$

where $\{h_j\}$ is a basis for the CS and f_0 is orthogonal to the regression function of every statistic, i.e.,

$$\int_I E(S|\theta = t) f_0(t)dt = 0$$

for every statistic S . In this decomposition, f_0 is called the null part of f , f^* the identifiable part of f and α_j

the j^{th} coordinate of f .

The identifiable part of f is indeed identifiable because a sequence of estimators $\{\hat{f}_N(t)\}$ can be constructed that will (almost surely) converge to $f^*(t)$ as sample size N is increased. The convergence turns out to be uniform in t .

The null part of f is null in the sense that it is totally unrelated to data. There is no objective way to use the administered items to distinguish two densities with the same identifiable parts and different null parts. Such densities cannot and, for most purposes, need not be distinguished. Both densities lead to the same predictions in all applications. A proposition or scientific statement that is true if f_1 is the ability density and false if f_2 is the ability density is, in the technical, foundations of measurement sense of the term, not meaningful. The proposition may be interesting, clearly stated and important, but there will be no way to tell if it is true or false from the observed responses to the test items.

The representation leads to a strategy for estimating densities. If the h_j (called coordinate functions) are orthonormal, then the coordinates α_j have a statistical interpretation,

$$\alpha_j = E[h_j(\theta)] ,$$

i.e., α_j is the expected value of a function of the unobserved ability θ .

Since h_j is in CS , h_j is the regression function of some statistic, say X_j , and its regression function, the conditional expected

value of X_j , will be equal to h_j

$$E(X_j|\theta = t) = h_j(t) .$$

Since X_j is simply a function of item scores, X_j can be computed for each examinee in a large sample of, say, N examinees to obtain a sample mean $\overline{X}_j$. By the law of large numbers the estimate $\hat{f}_N(t)$

$$\hat{f}_N(t) = \sum_{j=1}^{J} \overline{X}_j h_j(t)$$

will converge (in probability) to the identifiable part of f as the sample size N becomes large.

This estimate is especially well behaved and easy to study because the examinees are independently sampled. In fact, for sufficiently large sample size N the vector of sample averages

$$N^{\frac{1}{2}} \langle \overline{X}_1, \overline{X}_2, \ldots, \overline{X}_J \rangle$$

will be nearly multivariate normal with covariance matrix that, at least in some applications, can be regarded as known.

These loosely presented observations and definitions are collected, summarized and extended in the following outline.

AN INTRODUCTION TO
MULTILINEAR FORMULA SCORE THEORY

Section Three: An Outline of Some Basic Theory

The preceding section is now outlined for clarity and ease of future reference. Some additional assumptions and notation are introduced in this outline.

Basic Latent Trait Model & Notation (Simplest one-dimensional version)

Ω = $\{\omega\}$ = the probability space

= an infinite set of actual or conceivable examinees available for sampling and testing

θ = ability random variable. $\theta(\omega)$ is unobserved.

$f(\cdot)$ = the density for θ. Its support will always be assumed to be contained in an interval I. Except when noted, only continuous densities are considered.

I = a closed interval containing all abilities. $\int_I f(t)dt = 1$

n = number of test items

U = a random n-vector of item scores, the only observables

= $\langle u_1, u_2, \ldots, u_n \rangle$

u_i = item score random variable. $u_i(\omega)$ is either zero or one.

$P_i(\cdot)$ = item response function. $P_i(t) = E(u_i | \theta = t)$. Also called an item characteristic function. These functions are assumed to be continuous.

Local independence assumption: For any n-vector of zeros and ones

$$U^* = \langle u_1^*, u_2^*, \ldots, u_n^* \rangle$$

$$\mathrm{Prob}\{U = U^* | \theta = t\} = \prod_{i=1}^{n} \{u_i^* P_i(t) + (1 - u_i^*)[1 - P_i(t)]\}$$

Statistic: A number-valued function of the item scores.

Basic Formula Score Terminology

Regression Function: The regression function of a statistic S is the conditional expectation

$$R_S(t) = E[S|\theta = t] .$$

Canonical Space: The real vector space of all regression functions for a test. Abbreviated CS. A finite dimensional subspace of the vector space of all continuous functions defined on I. It can be shown to have dimension $\leq 2^n$.

J: The dimension of the canonical space. Discussed in Section 4.

$(\cdot,\cdot)$: Notation for the inner product used on the space of continuous functions defined on I. $(g,h) = \int_I g(t)h(t)dt$. Note $(h,f) = E[h(\theta)]$.

Coordinate Functions: An orthonormal basis for the CS of a test. Generally denoted $\{h_1, h_2, \ldots, h_J\}$.

α_j: The projection of the ability density on the j^{th} coordinate function, h_j. Called the j^{th} coordinate of f. Has statistical interpretation $\alpha_j = E[h_j(\theta)]$.

Ability Density Equivalence

$\chi_A(\cdot)$: The indicator function of the set A

$$\chi_A(t) = 0 \text{ if } t \text{ is not in } A$$
$$= 1 \text{ if } t \text{ is in } A$$

$P[\cdot;S,\cdot]$: Notation for the probability distribution of the statistic S

$P[A;S,f_1]$ is the probability that the statistic S is in the set A when f_1 is the density for θ .

$$P[A;S,f_1] = \int E[\chi_A(S)|\theta = t]f_1(t)dt$$

Equivalent Densities: Two densities f_1, f_2 are equivalent if for every statistic S

$$P[\cdot;S,f_1] = P[\cdot;S,f_2]$$

i.e., if the probability distribution of each statistic is the same when $f = f_1$ as when $f = f_2$.

Characterization of Equivalent Ability Densities: f_1 is equivalent to f_2 if and only if

$$(f_1,h_j) = (f_2,h_j) \quad \text{for } j = 1,2,\ldots,J$$

for any set of coordinate functions $\{h_j\}$.

Ability Density Decomposition

$g = g_0 + g^*$: Every density g on I can be expressed uniquely in the form

$$g(\cdot) = g_0(\cdot) + g^*(\cdot)$$

where g^* is in the canonical space and for every statistic S

$$(R_S, g) = (R_S, g^*)$$

g^*: The <u>identifiable part</u> of a density g. The projection of the density into the CS.

$$g^*(t) = \sum_{j=1}^{J} \alpha_j h_j(t)$$

for coordinate functions $h_1, h_2, \ldots h_J$.

g_o: The <u>null part</u> of a density g.

$$g_o = g - g^*.$$

Cannot and generally need not be estimated because $f_1^* = f_2^*$ implies $P[\cdot;S,f_1] = P[\cdot:S,f_2]$ for <u>every</u> statistic S.

<u>Ability Density Representation</u>

Densities and J vectors: The mapping

$$g \rightarrow <(g,h_1),(g,h_2),\ldots,(g,h_J)>$$

associates each density on I with a unique J vector. Densities associated with the same J vector are equivalent.

<u>Consistent, Unbiased Estimates of the Identifiable Part of the Ability Density</u>

X_j: A statistic with regression function equal to h_j; i.e., $R_{X_j}(\cdot) = h_j(\cdot)$. There must be at least one because

h_j is in the CS and the CS consists of regression functions only.

X_j must be bounded because it has all of its probability on a set of 2^n points.

$E(X_j) = \alpha_j$: Follows from

$$
\begin{aligned}
E(X_j) &= E[E(X_j|\theta)] \\
&= E[R_{X_j}(\theta)] \\
&= E[h_j(\theta)] \\
&= (h_j,f)
\end{aligned}
$$

$\overline{X}_{j,N}$: Sample mean of X_j from a sample of N examinees.

$\overline{\underset{\sim}{X}}_N = \langle\overline{X}_{1,N},\overline{X}_{2,N},\overline{X}_{J,N}\rangle$: Sample mean of N bounded independent, identically distributed random vectors. Converges to $\langle\alpha_1,\alpha_2,\ldots,\alpha_J\rangle$. Asymptotically $N^{\frac{1}{2}}\,\overline{\underset{\sim}{X}}_N$ is multivariate normal.

$\hat{f}_N(t)$: $\hat{f}_N(t) = \sum_{j=1}^{N} \overline{X}_{j,N} h_j(t)$

a consistent, unbiased estimate of the identifiable part of the ability density. Asymptotically $N^{\frac{1}{2}}\,\hat{f}_N(t)$ is normal.

Construction of Coordinate Functions h_j and Coordinate Estimators X_j: See Section 4.

AN INTRODUCTION TO
MULTILINEAR FORMULA SCORE THEORY

Section Four: Implementation of Ability Density Results

For some applications we have been able to improve upon CFSM (described below) by developing a parallel theory in which the item scores u_i are replaced by the scores $w_i=2u_i-1$. The "complete" set of statistics is the set of all products of these w_i *instead of* u_i. *The components of* CFSM were regression functions $R_{u_i}(\cdot)=P_i(\cdot)$, an easily computed formula for associating item response patterns with continuous functions, $V(t)=\Pi[1+u_iP_i(t)]$, *and an operator defined by the* function $H(s,t)=E[V(t)|\theta=s]=\Pi[1+P_i(s)P_i(t)]$. The analogous components for the new method are described at the end of this section.

The abstract results given in the preceding sections are being used now to estimate densities and item characteristic curves. This section is included to show in a general way how the theory is used to analyze data. The discussion is organized about four technical questions that commonly arise in response to presentations of the theory.

1. How are the coordinate functions $h_1, h_2, \ldots, h_J$ determined in actual applications?

2. How are the statistics X_j specified?

3. What is the dimension J and how is it determined?

4. Can the calculations be arranged in a way to avoid very long, numerically unstable calculations?

A set of statistics $\{S_k\}$ is called <u>complete</u> if any stati tic S can be written as a linear combination of finitely many of them. For example, the <u>elementary</u> formula scores $\{v_k\}$ formed by considering all products of item scores are complete. These are the scores

$$\begin{array}{l} 1 \\ u_1, u_2, \ldots, u_n \\ u_1 u_2, \ldots, u_{n-1} u_n \\ \vdots \\ u_1 u_2 \cdots u_n . \end{array}$$

The regression functions $R_{v_k}(t) = E[v_k|\theta = t]$ are simply products of item characteristic curves P_i.

A set of coordinate functions can be constructed from any complete set of statistics as follows. First a function on $I \times I$ is specified by the formula

$$H(s,t) = \sum_k R_{S_k}(s)\, R_{S_k}(t).$$

"Is specified" in practical terms means that a computer subroutine is written that accepts pairs of numbers s,t as input and returns the number given on the right hand side as output. In the special case where $\{S_k\} = \{v_k\}$, the "elementary" formula scores, it can be shown that H has the easily calculated form,

$$H(s,t) = \prod_{i=1}^{n} [1 + P_i(s)P_i(t)].$$

This special case will be discussed after the general case is treated.

Using standard methods $H(s,t)$ is decomposed into a finite sum of products of orthonormal functions. More specifically, a set of positive numbers

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_J > 0$$

and orthogonal functions h_j, $j = 1,2,\dots,J$

$$(h_j,h_{j'}) = \begin{cases} 1 & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases}$$

are computed such that

$$(*) \quad H(s,t) = \sum_{j=1}^{J} \lambda_j h_j(s) h_j(t)$$

for all s,t in I . Just as the eigenvalues of a positive semidefinite matrix are determined by the matrix, the λ's and the "rank" J are determined by H . It can be shown that every function in the CS is a linear combination of the h_j no matter which complete set of statistics is used to construct H . In other words, the h_j are coordinate functions, and J is the dimension of the CS .

To construct an estimator of the coordinate $\alpha_j = E[h_j(\theta)]$, note that h_j is an eigenfunction of the "linear operator" defined by H . In symbols,

$$\lambda_j h_j(t) = \int H(s,t)\, h_j(s)\, ds$$

$$= \sum_k R_{S_k}(t) \int R_{S_k}(s)\, h_j(s)\, ds$$

$$= \sum_k R_{S_k}(t)\, (R_{S_k}, h_j)\ .$$

If $R_{S_k}(t) = E[S_k|\theta = t]$ is replaced by S_k/λ_j in this formula, a statistic X_j is specified,

$$X_j = \lambda_j^{-1} \sum_k S_k(R_{S_k}, h_j).$$

Since $E(S_k|\theta = t)$ is $R_{S_k}(t)$ it follows that $E(X_j|\theta = t) = h_j(t)$ and $E(X_j) = E[h_j(\theta)]$.

Thus the sample mean $\overline{X}_j$ is a consistent, unbiased estimator of the coordinate α_j.

J, the dimensionality of the CS of the test, was calculated by analyzing any complete set of statistics $\{S_k\}$. J turned out to be the "rank" of the linear operator

$$h \rightarrow \psi(h) = \int \sum_k R_{S_k}(\cdot) R_{S_k}(t) h(t) dt = \int H(\cdot,t) h(t) dt\ .$$

Although the eigenfunctions h_j and the eigenvalues λ_j depend on the choice of $\{S_k\}$, J does not. However, in some applications we may be led to particular $\{S_k\}$ and subsequently to a decision to treat the CS as if it had dimension $J' < J$.

The typical situation arises when one considers the problem of selecting $\hat{f}$ so as to minimize the quadratic index of goodness of fit

$$Q(\hat{f}) = \sum_k [\overline{S}_k - E(S_k;\hat{f})]^2.$$

Here $\overline{S}_k$ is the sample average value of S_k and

$$E(S_k;\hat{f}) = \int_I R_{S_k}(t)\hat{f}(t)dt$$

is the predicted value of $\overline{S}_k$. If $H(s,t) = \sum_{j=1}^{J} \lambda_j h_j(s) h_j(t)$

then Q can be written in the form

$$(**) \quad Q(\hat{f}) = \sum_{j=1}^{J} \lambda_j [\hat{\alpha}_j - \overline{X}_j]^2 + \text{terms that are independent of } \hat{f}$$

where $\hat{\alpha}_j = (\hat{f},h_j)$ and $\overline{X}_j = \lambda_j^{-1} \sum_k \overline{S}_k (R_{S_k},h_j)$.

From formula (**) it can be seen that the size of λ_j measures the degree of improvement of fit to a set of statistics $\{S_k\}$ that can be obtained by including one more term in the representation of the identifiable part of the ability density

$$\sum_{j'<j} \alpha_{j'} h_{j'}(\cdot).$$

If λ_j is very small or $\overline{X}_j$ has very large sampling error, then we generally do not attempt to estimate the coordinate and proceed

as if the canonical space had lower dimensionality than J.

In applications we have been selecting J' by computing the λ_j and treating very small λ_j's as zero. As a check on the adequacy of this procedure we examine the difference between $\sum_{j=1}^{J'} (g,h_j)h_j(\cdot)$ and $g(\cdot)$ for various functions g. The functions g we generally consider are the P_i, selected regression functions and various guesses about $f(\cdot)$. The two functions will agree exactly if g is in the CS and J' is the dimensionality of the CS.

This section is concluded with a brief discussion of a formula scoring technique that has proven more powerful and accurate than all of the other techniques we have tried. The complete formula score method (CFSM) uses the elementary formula scores $\{v_k\}$ as its complete set of statistics and begins with the identity

$$H(s,t) = \sum_k R_{v_k}(s)R_{v_k}(t)$$

$$= \prod_{i=1}^{n} [1 + P_i(s)P_i(t)].$$

(The identity is verified by induction or by expanding the product.) The sum has 2^n terms, but the product has only n terms and thus can be accurately calculated with many fewer operations.

The X_j can also be calculated with a variant of this identity. Each sampled examinee's data is transformed to define a random continuous function $V(t)$, which is called the V-transform of the

examinee's data. This function

$$V(t) = \prod_{i=1}^{n} [1 + u_i P_i(t)]$$

is easily calculated and is identically equal to

$$\sum_{k=1}^{2^n} v_k R_{v_k}(t) \; .$$

Therefore $\int_I V(t)h_j(t)dt$ equals $\sum_k v_k(R_{v_k}, h_j) = \lambda_j X_j$. This reduces the calculation of X_j from 2^n operations to one numerical integration. In fact, to calculate the sample mean $\overline{X}_j$ only one integration need be done. This is true because the sample average of the integrated $V(t)$ is the integral of the sample average

$$\text{Ave}\{\int V(t)h_j(t)dt\} = \int [\text{Ave}\, V(t)]h_j(t)dt.$$

Thus in applications, $V(t)$ is computed on a grid of t values for each examinee, accumulated over examinees and numerically integrated to obtain the J sample means $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_J$. This procedure can be adapted to compute sample covariances of the X_j by numerical integration.

Continuation of Notes on Section Four

A Generalization of CFSM

Components of CFSM:

$$u_i$$

$$R_{u_i} = P_i$$

$$V(\cdot) = \Pi[1+u_i P_i(\cdot)]$$

$$H(s,t) = E[V(t)|\theta=s]$$

Analogues for LFSM:

$$w_i = 2u_i-1$$

$$R_{w_i}(\cdot) = 2P_i(\cdot)-1$$

$$L(\cdot) = \prod_{i=1}^{n} [1+w_i R_{w_i}(\cdot)]$$

$$G(s,t) = \prod_{i=1}^{n} [1+R_{w_i}(s) R_{w_i}(t)] .$$

Note that the likelihood function is proportional to L . Thus in applications requiring a representation of likelihood functions as linear combinations of a small number of coordinate functions, the eigenfunctions of G (or a modification of G used to avoid numerical problems) are now commonly used.

AN INTRODUCTION TO
MULTILINEAR FORMULA SCORE THEORY

Section Five: Discovering the Shapes of Item Response Functions

The "density ratio" estimation procedure on the following pages was improved by implementing the first two refinements at the end of the section. In the process of implementing the third, an alternative strategy was discovered. The new strategy has the generality of density ratio estimation, but it makes much more efficient use of data. The key idea is sketched at the end of this section.

Sometimes -- but not always -- the shape of item characteristic curves can be rationally deduced and parameterized. For example, the S-shape of the three-parameter logistic curve on an unbounded ability continuum follows from: (1) Monotonicity (more able examinees are more likely to answer correctly), (2) Asymptotes (although the probability of a correct response can be made arbitrarily close to one by sampling from very high ability subpopulations, a substantial proportion of each low-ability subpopulation will select the correct option of a well-constructed multiple choice item.) (3) Simplicity/Parsimony (the item response curve has no more points of inflection than the one implied by (1), (2) and smoothness conditions.) (4) Symmetry (the graph of the item characteristic curve is symmetric for high and low ability levels in the sense that a length-preserving transformation $(x,y) \rightarrow (2x_0-x, 2y_0-y)$ about a point of inflection on the graph (x_0,y_0) carries the curve into itself.)

Fitting three parameter logistic functions is sensible when these conditions are met because every curve satisfying these conditions will be close to at least one three-parameter logistic function.

Sometimes these assumptions are implausible or clearly false, and a method is needed to discover and parameterize shape. As a one-dimensional example, it was demonstrated (Levine and Drasgow, 1983) with a very large sample of aptitude test examinees that the conditional response function for the response of choosing a particular (wrong) option on a multiple choice test for several items has this bow shape: .

In multidimensional measurement, the shape of the item characteristic surface is a matter of considerable psychological importance because it represents a statement about how several abilities interact to simultaneously determine response probability. In the next section, the methods discussed in this section are used to develop a procedure for determining the shape of an item characteristic surface.

Another application of the method described in this section is the study of item responses, such as omitting, that cannot reasonably be expected to satisfy local independence. FST permits one to construct a consistent estimate of an "omitting characteristic curve"

$$P\{\text{item } i \text{ is omitted} \mid \theta = t\}$$

without assuming local independence for omitting responses.

The basic issues addressed in this segment of our research are

1. What is the shape of a new item or item type's item characteristic curve?

2. What information about ability is contained in wrong answers and item omitting?

3. Model item responses that may not be locally independent such as omitting and also the following examples drawn from computer administered tests:

   (i) attempting to change an answer

   (ii) requesting a display of a previously presented item or part of an item

(iii) responding in a time clearly too short to read the item

<u>Equations Relating FST to ICC Shape</u>: Our results in this area depend on the following elementary relation which is used to reduce item response function estimation to ability density estimation:

$$\text{Prob}\{\text{correct response} \mid \text{ability is in set } A\} = \frac{\text{Prob}\{\text{ability is in } A \mid \text{correct response}\}}{\text{Prob}\{\text{ability is in } A\}} \times \text{Prob}\{\text{correct response}\}$$

Thus the conditional response probability is proportional to the ratio of the ability distribution in the item-passing subpopulation to the unconditional distribution. (The constant of proportionality, Prob{correct response} = $\overline{P}$, has an obvious consistent unbiased estimate, the sample proportion correct.) If regularity assumptions are made then

$$P(t) = \frac{f^{+}(t)}{f(t)}\overline{P}$$

where f^{+} = the ability density in the subpopulation of item-passers

f = the ability density.

This equation can be written in the form

$$P(t) = \frac{1}{1 + \frac{\overline{Q}}{\overline{P}}\,\frac{f^{-}(t)}{f^{+}(t)}}$$

where $\bar{Q} = 1 - \bar{P}$ and f^- is the failure density.

This formula is especially useful for developing a distribution theory for the estimates because the failure and passing subpopulations are disjoint. Thus the distribution of an estimator of $f^+(t)/f^-(t)$ can be developed by studying the ratio of statistically independent random variables.

Several variations of this approach to item characteristic curve estimation have been tried with generally satisfactory but occassionally poor results. Systematic comparisons of the variations will be made after more exploratory work. Some current and projected refinements are sketched below.

(1) If ability is uniformly distributed or if $f(\cdot)$ is constant on a range of abilities of interest, then the ICC is proportional to f^+ and ICC estimation is simplified. Furthermore, sampling fluctuation is unlikely to give a zero or negative density estimate. An initial approximation can be used to transform ability so that $f(\cdot)$ is approximately constant.

(2) The sampling distribution of coordinate estimates depends on the choice of $\{S_k\}$. We have observed that if h_1, the coordinate function with the largest eigenvalue is close to $f(\cdot)$ then very good estimates of $f(\cdot)$ are obtained. Currently an attempt is being made to capitalize on this effect by carefully choosing $\{S_k\}$ and controlling the eigenfunction shapes.

(3) The current density estimates are least squares in the sense that they minimize the residual sum of squares Q (Section 4). The results on density equivalence permit the expression of the likelihood function as a function of finitely many parameters, the coordinates α_j. In principle, one could maximize this expression and compute maximum likelihood density estimates.

The basic equation used to relate item characteristic curves to density ratios is essentially the definition of conditional probability. Local independence plays no role in its derivation. In fact the binary item response on the focal item is being used merely to divide the sample into item passers and item failers. The equation would remain valid if any binary score were used to dichotomize the sample and population. Therefore the same estimation techniques used to estimate item characteristic curves could be used to study the relation of ability to complex item responses that failed to satisfy the local independence assumption. These include item skipping and very fast responding.

Continuation of Notes on Section Five

Maximum Likelihood FST Item Response Curve Estimation

If $P(t) = \sum_{j=1}^{J} \alpha_j^+ h_j(t)$ is the item response function for item score u and

$$\ell(U_a^*;\ t) = \text{Prob}\{U=U_a^* | \theta=t\}$$

is the likelihood function for examinee a's response pattern U_a^* on calibrated items, and f is the ability density, then

$$\begin{aligned}
&\text{Prob}\{U=U_a^* \ \&\ u=u_a^*\} \\
&\quad = \int\{(u_a^* P(t) + (1-u_a^*)[1-P(t)]\}\ell(U_a^*;\ t)f(t)dt \\
&\quad = \int\{(2u_a^*-1)\Sigma\alpha_j^+ h_j(t) + u_a^*\}\ell(U_a^*;\ t)f(t)dt \\
&\quad = \{u_a^* \int \ell(U_a^*;\ t)f(t)dt\} + \sum_j \alpha_j^+ (2u_a^*-1)\int h_j(t)\ell(U_a^*;\ t)f(t)dt
\end{aligned}$$

is a linear function of the unknown α_j^+. Thus the log likelihood function $\sum_a \log \text{Prob}\{U=U_a^* \ \&\ u=u_a^*\}$ is convex.

Furthermore, the set of vectors $\{\underset{\sim}{\alpha}^+ : 0<\Sigma\alpha_j^+ h_j(t)<1\}$ is convex. Thus maximum likelihood coordinate estimates can be computed by solving a standard mathematical problem: maximize a convex function on a convex set.

Maximum likelihood is now being used to estimate item response functions and ability distributions.

AN INTRODUCTION TO MULTILINEAR FORMULA SCORE THEORY

Section Six: Multidimensional Formula Scoring for Homogeneous Subtests

The key idea in this section is the representation of the canonical space of two tests as the tensor product of the canonical spaces for each test. With this representation the application of one dimensional results to estimate multivariate distributions is straightforward. In particular, previously discussed "density ratios" or maximum likelihood coordinate estimates can be used to estimate bivariate item response surfaces. It should have been emphasized that the multivariate ability distributions are intrinsicly important.

There is an interesting and important multidimensional measurement problem that can be implemented with currently available, unidimensional software. The problem is to discover how several abilities jointly determine response probability on new item types. This can be done when certain psychological assumptions are valid. In particular, it is necessary that a variety of item types are available, that all items depend on the same small number of abilities and that items can be grouped into "homogeneous" subtests.

For concreteness consider three item types: synonyms, antonyms and analogies. Suppose that all three depend only on a pair of abilities θ_1, θ_2 in the sense that for any ability levels s,t and any r items, the item scores $u_{i_1}, u_{i_2}, \ldots, u_{i_r}$ satisfy

$$E[\prod_{j=1}^{r} u_{i_j} | \theta_1 = s \;\&\; \theta_2 = t] = \prod_{j=1}^{r} E[u_{i_j} | \theta_1 = s \;\&\; \theta_2 = t] \cdot$$

In other words the items are independent in subpopulations formed by conditioning on both abilities.

Homogenous subtests, defined more formally below, are unidimensional subtests of a multidimensional test. For example both synonym and antonym items may require both language fluency θ_1 and an ability to recognize and generalize abstract relations θ_2. But the antonym items can be written in such a way as to demand a relatively large amount of the second ability. Thus synonym items may satisfy local independence with respect

to some linear or nonlinear function of θ_1 and θ_2, say $\theta_1 + \theta_2$ and antonym items with respect to, say $\theta_1 + 2\theta_2$. Subtests consisting of one item type only will appear unidimensional, but the total test will not.

Note that the assumption that item types form homogeneous subtests is more general (and more believable) than the assumption that different item types measure different traits. This should be obvious after the discussion of "ad hoc coordinates."

If certain plausible assumptions (written out below) are correct then off-the-shelf unidimensional parameter estimation programs can be used with a test consisting of only synonym items and a test consisting of only antonym items. FST can then be used to represent and estimate bivariate analogy item response functions. After some preliminary definitions, additional details are given.

Homogenous subtests and ad hoc coordinates:

Latent trait theory provides a way to precisely state what is meant by a homogeneous subtest and items requiring different amounts of unobserved abilities. As before the population of examinees is denoted by a point set $\Omega = \{\omega\}$. In this situation, abilities map examinees into 2-vectors rather than numbers: $\theta(\omega) = \langle\theta_1(\omega), \theta_2(\omega)\rangle$. Because examinees are randomly sampled θ is a random vector.

To quantify the notion of homogeneous subtests, a pair of number valued functions ϕ, ψ are considered. The first subtest is homogeneous in the sense that it satisfies a local independence assumption relative to $\phi(\theta)$.

In symbols, for items $i_1, i_2, \ldots, i_r$ on the first subtest

$$\text{Prob}\{u_{i_1} = 1, u_{i_2} = 1, \ldots \text{ and } u_{i_r} = 1 | \phi(\theta) = t\} = \prod_{j=1}^{r} P_{i_j}(t)$$

where $\quad P_{i_j}(t) = \text{Prob}\{u_{i_j} = 1 | \phi(\theta) = t\}$.

A similar equation expresses the assumption that the second subtest is homogeneous with respect to $\psi(\theta)$: The item scores for items in the second subtest are independent in the subpopulation of examinees with the property $\psi(\theta) = s$ for each constant s .

Note that these conditions permit using available parameter estimation programs to calibrate each subtest separately. Somewhat paradoxically, each item is essentially multidimensional but each subtest satisfies the axioms of one dimensional latent trait theory. FST provides a method for integrating the subtests and modelling the analogy items that also depend on fluency and abstraction, but to an unknown extent.

The functions ϕ and ψ are called "ad hoc coordinates." If for each $\langle s,t \rangle$ there is at most one vector $\langle x,y \rangle$ satisfying

$$s = \phi(x,y)$$

$$t = \psi(x,y)$$

and certain regularity conditions are met then ϕ and ψ can be treated as curvilinear coordinates for the set of (bivariate) abilities.

The FST analysis to be presented only gives a representation of item response function in terms of ad hoc coordinates

$$P_i(s,t) = \text{Prob}\{u_i = 1 | \phi(\theta) = s \;\&\; \psi(\theta) = t\}$$

For many modelling problems this is adequate. Admittedly the variables θ_1 (fluency) and θ_2 (abstraction) are considerably more interesting than ϕ and ψ. Conjoint measurement or uniform systems analysis (Levine, 1970) may permit one to analyze the relation between ϕ,ψ and θ_1,θ_2. However, our current concern is with the psychometric problem of representing new items in the ad hoc coordinate system.

Formula Score Approach to Representing New Item Types

Consider a twenty-one item test consisting of 10 synonym items followed by 10 antonym items and one analogy item. Our task is to compute

$$\mathrm{Prob}\{u_{21} = 1 | \phi(\theta) = s \;\&\; \psi(\theta) = t\} = P(s,t) .$$

We propose to first analyze the homogeneous subtests separately. The most direct approach would be to first calibrate the synonym items by embedding them in a large conventional administration of many items of the same type. The analysis would yield ϕ item characteristic curves

$$\mathrm{Prob}\{u_i = 1 | \phi(\theta_1,\theta_2) = t\} = P_i(t) \qquad i = 1,2,\ldots 10 .$$

Similarly a separate analysis of the antonym items yields ψ item characteristic curves

$$\mathrm{Prob}\{u_i = 1 | \psi(\theta_1,\theta_2) = t\} = P_i(t) \qquad i = 11,12,\ldots,20 .$$

To discover the shape of an analogy item's ICC, a 21 item test would be administered to a sample of examinees. By an obvious generalization

in Section 5, the analogy item ICC

$$P(s,t) = \text{Prob}\{u_{21} = 1 \mid \phi(\theta_1,\theta_2) = s \text{ and } \psi(\theta_1,\theta_2) = t\}$$

can be represented as a ratio of densities

$$P(s,t) = \frac{f^+(s,t)}{f\ (s,t)} \bar{P}$$

where f^+ is the (bivariate) density in the subpopulation of item 21 passers, f is the unconditional density.

Before continuing this analysis, the CS for the 20 item test is described. It turns out that the assumptions imply that the canonical space of the twenty item test is simply the "tensor product" of the CS for the first subtest and the second subtest.

At this point it seems advisable to restate some definitions. The CS for the first subtest is the set of one-dimensional regression functions

$$E[S \mid \phi(\theta_1,\theta_2) = t] = R_S(t)$$

where S is a statistic whose value depends on the first ten item scores only. The CS for the antonym item is the set of regression functions $E[S \mid \Psi(\theta_1,\theta_2) = t]$ for statistics S that are functions of the second ten scores only.

The CS for the first twenty item subtest will be the set of regression functions

$$R_S(s,t) = E(S \mid \phi(\theta_1,\theta_2) = s \text{ and } \psi(\theta_1,\theta_2) = t)$$

for the statistics S of the first 20 scores. Our psychometric assumptions imply that any $R(s,t)$ in the 20 item CS can be written as a finite sum

of functions of the form $h(s)h'(t)$ for h in the first subtest CS and h' in the second subtest CS. In fact if $\{h_1, h_2, \ldots, h_J\}$ is a basis for the first CS and $\{h_1', h_2', \ldots, h_{J'}'\}$ for the second CS then the $J \times J'$ functions

$$h_j(s)h_{j'}'(t) = h_{jj'}(s,t)$$

$$j = 1,2,\ldots,J$$

$$j' = 1,2,\ldots,J'$$

will be a basis for the two-dimensional CS.

Current one dimensional FS programs can be used to estimate the bivariate densities f^+ and f. The estimate will have form

$$f^+(s,t) = \sum_{j=1}^{J} \sum_{j'=1}^{J'} \overline{X}_{jj'} h_{jj'}(s,t)$$

where the sample mean $\overline{X}_{jj'}$ is a consistent, unbiased estimator of $E[h_{jj'}(\theta)]$. If f^+ and f are in the 20 item CS, then (by sample splitting to estimate f^+ and f separately) a consistent estimate of $P(s,t)$ is easily specified, and the shape of the item response surface can be "discovered."

The success of this approach requires f^+ and f to be in or near the 20 item, two-dimensional CS. This assumption seems plausible when one considers the variety of shapes that can be constructed as linear combinations of the $J \times J'$ coordinate functions.

AN INTRODUCTION TO
MULTILINEAR FORMULA SCORE THEORY

Section Seven: Are Two Tests Measuring the Same Trait(s)?

A colleague (Fritz Drasgow) has convinced me that I am asking the wrong question in this section. Different tests generally measure different traits; the important question is, "How different are the traits?" If two abilities measured by a pair of tests are equal, then the probability of sampling an examinee with $\theta_1 \neq \theta_2$ will be zero. Therefore we have been attempting to quantify the difference between abilities by using the results in Section Six to estimate bivariate distributions and the expectations of functions of bivariate abilities such as $E[(\theta_1-\theta_2)^2]$.

Suppose a major revision is made of a complex, not necessarily unidimensional test. Does the new test measure the same traits? Suppose the format or mode of administration is changed. Does the test still measure the same traits? Suppose a translation of the test is attempted into another language and that it is unlikely that every original test item is psychometrically equivalent to its translation. Can the translated test nonetheless measure the same traits as the original? These questions have led us to ask the foundations of measurement question,

> What necessary and/or sufficient conditions must item scores obey before it can be concluded that two nonparallel tests are measuring the same trait(s)?

It is hoped that theoretical work on this problem *will lead* a statistical test that can be used in applications. This section relates FST to the problem and reports some current work.

Two nonparallel tests are administered to the same population. (The tests can be thought of as subtests of one test.) Item response curves are fitted to each test separately. In otherwords, except possibly to compute an equating transformation, only test-one scores are used to estimate the IRF of a test-one item. Can the variable in the first set of IRF's be given the same interpretation as the variable in the second set of IRF's ?

The mathematical kernel of this problem seems to be this. A probability measure is given for a set Ω along with two sets of zero-one random variables

$$u_1, u_2, \ldots u_n$$

$$u_1', u_2', \ldots u_{n'}'$$

Two sets of real functions

$$P_1, P_2, \ldots P_n$$

$$P'_1, P'_2, \ldots P'_{n'}$$

are also given. What conditions must be assumed in order for it to be possible to construct one more random variable θ such that the P's are item response functions for the u's relative to θ and all $n + n'$ u's are locally independent relative to θ .

A strong necessary condition on the given scores and functions can be formulated with FS notation and concepts. Let $\{v_k\}$ denote the elementary scores in the first set of u's (Section 4). Let $\{v'_{k'}\}$ denote the elementary scores in the second set of u's . Let R_{v_k} and $R_{v'_{k'}}$ be the corresponding products of P's . Thus if

$$v_k = u_{i_1} u_{i_2} \ldots u_{i_r}$$

then $R_{v_k}(\cdot) = P_{i_1}(\cdot) P_{i_2}(\cdot) \ldots P_{i_r}(\cdot)$.

Let $\{v_{k,k'}\}$ denote the elementary formula for the $n + n'$ item test where $v_{k,k'}$ is $v_k v'_{k'}$. Let $R_{k,k'}$ denote the corresponding product of functions for $v_{k,k'}$, i.e., $R_{k,k'} = R_{v_k} R_{v'_{k'}}$. And let $\{h_j\}_{j=1}^{J}$ be an orthonormal basis for the linear span of the $\{R_{k,k'}\}$. For example, the h_j could be obtained by analyzing the function H(s,t) defined by

$$\prod_{i=1}^{n} [1 + P_i(s)P_i(t)] \prod_{i=1}^{n'} [1 + P'_i(s)P'_i(t)]$$

as in Section 4. Finally, let X_j be the random variable

$$X_j = \sum_k \sum_{k'} v_{k,k'} (R_{k,k'}, h_j) ,$$

which can be more conveniently computed as described in Section 4.

It is easy to show the following condition on expected values of scores is necessary

$$(1) \quad E(v_k v_{k'}) = \sum_{j=1}^{J} E(X_j)(h_j, R_{k,k'})$$

for all elementary formula scores $v_k, v_{k'}$.

The condition also appears to be sufficient, although a proof is not on hand at this time. In any event, additional work is needed to determine when one random variable suffices for Ω and specifed subpopulations of Ω .

Acknowledgements

This research was supported by Office of Naval Research Contracts N00014-79C-0752, NR 154-445 and N00014-83K-0397, NR 150-518.

References

Johnson, N. L. & Kotz, S. Continuous univariate distributions (Vol. 2). Boston: Houghton-Mifflin, 1970.

Levine, M. V. Transformations that render curves parallel. Journal of Mathematical Psychology, 1970, 7, 410-443.

Levine, M. V. Identifying different item response curves. Air Force Human Resources Laboratory Report TP 83-88. San Antonio Texas: Air Force Systems Command, 1983.

Levine, M.V., & Drasgow, F. The relation between option choice and estimated ability. Educational and Psychological Measurement, 1983, 43, 675-685.

Williams, B. Distributions of central quadratic forms in normal variables. Model-Based Measurement Laboratory Report 84-3. University of Illinois, Department of Educational Psychology, 1983.

Navy

1 Dr. Nick Bond
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

1 Dr. Robert Breaux
NAVTRAEQUIPCEN
Code N-095R
Orlando, FL 32813

1 Dr. Robert Carroll
NAVOP 115
Washington , DC 20370

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Dr. John Ellis
Navy Personnel R&D Center
San Diego, CA 92252

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Mr. Dick Hoshaw
NAVOP-135
Arlington Annex
Room 2834
Washington , DC 20350

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

1 Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

1 Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Richard Snow
Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Edward Wegman
Office of Naval Research (Code 411S&P)
800 North Quincy Street
Arlington, VA 22217

Navy

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 Jerry Lehnus
CAT Project Office
HQ Marine Corps
Washington , DC 20380

1 Headquarters, U. S. Marine Corps
Code MFI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Ccde MPI-20)
Washington, DC 20380

Army

1 Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Blvd.
Alexandria , VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commander, U.S. Army Research Institute
for the Behavioral & Social Sciences
ATTN: PERI-BR (Dr. Judith Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joyce Shields
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Air Force

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. Patrick Kyllonen
AFHRL/MOE

Brooks AFB, TX 78235

1 Lt. Col. Dave Payne
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Roger Pennell
Air Force Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Major John Welsh
AFHRL/MOAN
Brooks AFB , TX 78223

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defens
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. W. Steve Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 Dr. Robert A. Wisher
OUSDRE (ELS)
The Pentagon, Room 3D129
Washington, DC 20301

Civilian Agencies

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 32605

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52243

1 Dr. Glenn Bryan
6208 Poe Road
Bethesda, MD 20817

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

Private Sector

1 Dr. Walter Cunningham
University of Miami
Department of Psychology
Gainesville, FL 32611

1 Dr. Dattpradad Divgi
Syracuse University
Department of Psychology
Syracuse, NE 33210

1 Dr. Emmanuel Donchin
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Hei-Ki Dong
Ball Foundation
Room 314, Building B
800 Roosevelt Road
Glen Ellyn, IL 60137

1 Dr. Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel St.
Champaign, IL 61820

1 Dr. Susan Embertson
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Private Sector

1 Professor Donald Fitzgerald
University of New England
Armidale, New Scuth Wales 2351
AUSTRALIA

1 Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01002

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
603 East Daniel Street
Champaign, IL 61820

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Private Sector

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Advanced Statistical Technologies
Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

1 Dr. Marcel Just
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Demetrios Karis
Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820

1 Professor John A. Keats
Department of Psychology
The University of Newcastle
N.S.W. 2308
AUSTRALIA

1 Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Private Sector

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Mr. Phillip Livingston
Systems and Applied Sciences Corporatio
6811 Kenilworth Avenue
Riverdale, MD 20840

1 Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

1 Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Don Lyon
P. O. Box 44
Higley , AZ 85236

1 Dr. Scott Maxwell
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Mr. Robert McKinley
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Private Sector

1 Dr. Barbara Means
Human Resources Research Organization
300 North Washington
Alexandria, VA 22314

1 Dr. Robert Mislevy
711 Illinois Street
Geneva, IL 60134

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

1 Dr. Donald A Norman
Cognitive Science, C-015
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. James Olson
WICAT, Inc.
1875 Scuth State Street
Orem, UT 84057

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara , CA 93106

1 Dr. Steven E. Poltrock
Bell Laboratories 2D-444
600 Mountain Ave.
Murray Hill, NJ 07974

1 Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Private Sector

1 Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P. O. Box 688
Richardson, TX 75080

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

1 Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

1 Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

1 Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Private Sector

1 Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

1 Dr. V. R. R. Uppuluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

Private Sector

1 Dr. David Vale
Assessment Systems Corporation
2233 University Avenue
Suite 310
St. Paul, MN 55114

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin--Milwaukee
Milwaukee, WI 53201

1 Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90007

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Ms. Marilyn Wingersky
Educational Testing Service
Princeton, NJ 08541

1 Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

END

FILMED

10-84

DTIC